Received: 20 December 2019 | Accepted: 12 May 2020 | Published: 12 July 2020
DOI:10.20870/oeno-one.2020.54.3.2972

Using Bayesian growth models to predict grape yield

Rory Ellis[1],*, Elena Moltchanova[1], Daniel Gerhard[1], Mike Trought[2] and LinLin Yang[2]

[1]Department of Mathematics and Statistics, Erskine Building, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand
[2]Plant & Food Research 85 Budge Street, Blenheim, 7201, New Zealand

*Corresponding author: rory.ellis@pg.canterbury.ac.nz

ABSTRACT

Background and aims: Seasonal differences in vine yield need to be managed to ensure appropriate fruit composition at harvest. Differences in yield are the result of changes in vine management (e.g., the number of nodes retained after harvest) and weather conditions (in particular, temperature) at key vine development stages. Early yield prediction enables growers to manage vines to achieve target yields and prepare the required infrastructure for the harvest.
Methods and results: Bunch mass data was collected during the 2016/17, 2017/18 and 2018/19 seasons from a commercial vineyard on the Wairau Plains, Marlborough, New Zealand (41°2'23"S; 173°51'15"E). A Bayesian growth model, assuming a double sigmoidal curve, was used to predict the yield at the end of each season. The accuracy of the prediction was investigated using the Monte-Carlo simulation for yield prediction at various growth stages assuming different prior information.
Conclusion: The results show that the model is sensitive to prior assumption and that having a non-informative prior may be more beneficial than having an informative prior based on one unusual year.

KEYWORDS

Bayesian modelling, yield modelling, grapes, yield variability

INTRODUCTION

Seasonal differences in vine yield need to be managed to ensure appropriate fruit composition at harvest. Excessive yields can slow fruit development and result in infrastructure problems (e.g., inadequate tank space) at harvest. Weather conditions at flowering can cause significant differences in the percentage of flowers setting berries, hence affecting bunch mass (Trought, 2005). The ability to predict harvest bunch mass shortly after flowering is an essential tool for predicting vineyard grape yield. Grape berries exhibit a double sigmoid growth curve (Coombe, 1976). The initial phase (Phase I) is predominantly the result of cell division and expansion, which is followed by a slowing down in growth (fresh mass) phase (Phase II), in which the seed matures. This slow growth ends at veraison, with the onset of another period of berry growth, cell enlargement and metabolic changes (e.g., sugar accumulation and reduction in acidity (Phase III) (Tello and Forneck, 2018).

Sigmoid curves are often used to model various aspects of biological growth (Archontoulis and Miguez, 2015). The double sigmoidal curve, which combines two individual sigmoidal curves, is often used to model crop growth. Salisbury and Ross (2000) (cited by Parra-Coronado *et al.* (2016)) stated that fruits, such as raspberries, grapes, blackberries, olives and stone fruit, exhibit double sigmoidal growth behaviour. Galassi *et al.* (2000) produced curves for peach diameter and mass for cultivars in Ferrara and Rome in Italy. Fernandes *et al.*. (2017) compared various specifications of single and double logistic and Gompertz curves to determine which produced the highest adjusted-R^2values, concluding that the double sigmoidal curve (rather than the single sigmoids) was the optimal one. Parra-Coronado *et al.* (2016) specified a mathematical model for pineapple mass which takes into account altitude, as well as Growing Degree Days.

While Tarara *et al.* (2009) used a single sigmoid curve to monitor the response of grape mass per shoot and fruit to vegetative mass per shoot with respect to Growing Degree Days over five growing seasons, Coombe (1976) provided an intuitive rationale behind the use of the double sigmoidal curve for modelling the development of fleshy fruits, including grapes. The proposed curve incorporates periods of rapid growth intersected by a period of minimal growth and reflects the various stages of berry ripening well (Coombe and McCarthy, 2000), as illustrated in Figure 1. Phenologically, three stages are involved (Coombe, 1992). The first phase involves fruit set, when the grape inflorescences turn into grape berries; a period of rapid grape berry growth can be observed as cell division occurs (Parker, 2012). The second stage involves a lag phase, during which the berry growth slows and seed fresh mass reaches a maximum. The end of this stage (veraison) is followed by ripening (Trought, 2005), as sugars and water accumulate inside the grape berries, coinciding with a second period of rapid growth until, ultimately, the final yield is reached. The model described above can be implemented in various ways. Here, the use of a Bayesian analytic framework is proposed. Bayesian inference is becoming more popular as it is flexible in terms of model formulation, allows for easy incorporation of expert opinion via prior distributions, and, being based on Bayes' formula, can be easily updated continuously. The latter aspect is especially useful, enabling incorporation of additional information into the model to produce predictions, which not only reflect the initial conditions, but also management decisions and natural factors experienced during the growing season. The Bayesian framework is also good at handling missing data. This means that predictions can be

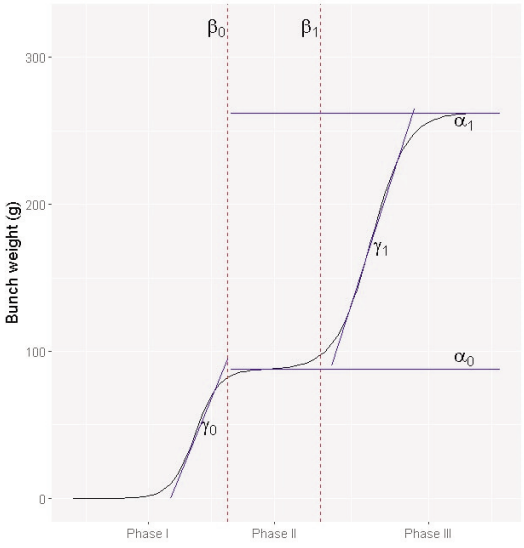

FIGURE 1. The parameterisation of the double sigmoidal growth curve, and the three main phases of phenological growth exhibited by grapes during the growing season.

made without requiring all of the information sources to be present at any particular time point. Should particular kinds of information not be available, the Bayesian model is simply able to integrate these missing components, updating itself instead using the information which is provided (Gelman *et al.*, 2014, Chapter 18).

When placing modelling within a dynamic context it is imperative to consider the Value of Information (VoI) aspect. Intuitively, the more data that is available, and the later into the season it has been collected, the more accurate the predictions will be for the end of the growing season. The choice between making a decision early on based on the information available or waiting for better information while losing some opportunities is the subject of economic theory, including the real options theory (Mun, 2002). While placing the motivation of updating the Bayesian model with data within strict economic theoretical context is beyond the scope of this paper, we examine how adding further time points to the data will affect the accuracy of the estimation.

We hereafter illustrate how the Bayesian framework is applied in order to fit a double sigmoid growth model to the grape bunch mass data collected from a commercial vineyard in New Zealand Marlborough region (41°2'23"S; 173°51'15"E) over three growing seasons. While doing so, the sensitivity of the model to prior information is examined, using both the observed and simulated data, alongside the Value of Information (VoI) aspect of the dynamic modelling. We conclude by summarising our observations and listing the most promising directions for further work in this area.

MATERIALS AND METHODS

Data on grape bunch masses for the 2016/17, 2017/18, and 2018/19 seasons were collected from a commercial vineyard on the Wairau Plains (41°2'23"S; 173°51'15"E). For the 2016/17 and 2017/18 growing seasons, fifteen replicate plots were established in a single vineyard row. Each plot consisted of four vines planted 1.8 m apart. A shoot with two bunches was randomly selected from each plot on approximately a weekly basis, starting at flowering and continuing until shortly before the commercial harvest. In the 2018/19 the number of plots was increased to 30 and sampling continued as before. Bunches were individually bagged and taken to the laboratory for weighing. There were a total of one, five, and 63 missing data points for the seasons 2016/17, 2017/18, and 2018/19 respectively.

2.1 Bayesian Inference

Before describing the details of the model, we would like to quickly review the basics of Bayesian Inference. Consider the data y and the likelihood g(y) and the prior distribution for the parameter of interest g(θ), which describes our understanding of the probable parameter values before the data are observed. The prior distribution incorporates the prior information available to the researcher and may be based on the general understanding of the phenomenon or on prior experience and expert opinion. When no information is available, the prior distributions tend to be very wide and are often called *vague*. Because the choice of prior may be deemed somewhat subjective, a sensitivity analysis is often performed to determine to what extent it affects the modelling result. The posterior distribution for the parameter may be obtained from the Bayes Theorem as follows:

$$g(\theta|y) = \frac{g(y|\theta)g(\theta)}{\int_y g(y|\theta)g(\theta)dy} \propto g(y|\theta)g(\theta)$$

Because the above derivation can rarely be done analytically, computationally intensive Markov Chain Monte Carlo (MCMC) methods are used to produce samples from the posterior distributions, for which summary statistics (such as posterior mean and posterior 95 % credible intervals) can then be obtained.

If the data y can be divided into two batches, y_1 and y_2, the equation above can be rewritten as

$$g(\theta|y_1, y_2) \propto g(y_1, y_2|\theta)g(\theta) \propto g(y_2|\theta)g(y_1|\theta)g(\theta) = g(y_2|\theta)g(\theta|y_1).$$

The posterior distribution of θ after observing the first batch y_1 thus becomes the prior distribution for the experiment involving y_2, leading to sequential updating.

A posterior predictive distribution for a new (future) observation $\tilde{y}$ given the data y can then be obtained as

$$g(\tilde{y}|y) = \int_y g(\tilde{y}|\theta)g(\theta|y)d\theta.$$

Again, numeric methods are customarily used to obtain a sample from such a distribution, which can then be summarised via sample statistics

such as mean and quantiles. We now turn to the double-sigmoidal growth model.

2.2 Double sigmoidal growth model

Let Yi describe the bunch mass observed at time xi , i = 1...n, where n is the total number of observations. In order to guarantee a non-negative response and control for heteroscedasticity apparent from Figures 1a, 1b, and 1c, it is common to assume that the logarithm of mass has a Gaussian distribution:

$$log(Y_i) \sim N(\mu_i, \tau) \qquad (1)$$

where τ is the *precision* or inverse variance, and the mean expected log-mass μ_i can be modelled via the double logistic curve as follows:

$$\mu_i = f(x_i, \alpha_0, \alpha_1, \beta_0, \beta_1, \gamma_0, \gamma_1) = \frac{\alpha_0}{1+e^{-\gamma_0(t_i-\beta_0)}} + \frac{\alpha_1}{1+e^{-\gamma_1(t_i-\beta_1)}} + \varepsilon_i \qquad (2)$$

where α_0 and α_1 describe the first and second asymptotes respectively, β0 and β1 are the inflection points, and γ_0 and γ_1 are the slope parameters. Figure 1 illustrates the role of these parameters further. In order to analyse the model within the Bayesian framework, we need to provide prior distributions for these parameters. It should be noted that we expect both slopes γ0 and γ1 to be positive. We also do not expect the second asymptote to be less than the first; i.e., $\alpha_1 \geq \alpha_0$. Note that when $\alpha_1 = \alpha_0$, the second term in Equation 2 disappears and the double sigmoidal curve collapses to a single sigmoidal curve. In order to reflect these restrictions, the following priors are assumed:

$$\alpha_0 \sim N(\mu_{\alpha 0}, \tau_{\alpha 0}) \qquad (3)$$

$$\alpha_1 \sim TN(\mu_{\alpha 1}, \tau_{\alpha 1}, \alpha_0) \qquad (4)$$

$$\beta_0 \sim N(\mu_{\beta 0}, \tau_{\beta 0}) \qquad (5)$$

$$\beta_1 \sim TN(\mu_{\beta 1}, \tau_1, \beta_0) \qquad (6)$$

$$\gamma_0 \sim TN(\mu_{\gamma 0}, \tau_{\gamma 0}, 0) \qquad (7)$$

$$\gamma_1 \sim TN(\mu_{\gamma 1}, \tau_{\gamma 1}, 0) \qquad (8)$$

$$\tau \sim Gamma(a_\tau, b_\tau) \qquad (9)$$

Where TN(μ, τ, L) denote a normal distribution with mean μ and precision τ left-truncated at L. In order to have *a priori* independent parameters, the model can be reparametrized in terms of $\Delta\alpha = \alpha_1 - \alpha_0$ and $\Delta\beta = \beta_1 - \beta_0$, and the priors in Eq. (4) and Eq. (6) recast as follows:

$$\Delta\alpha \sim N(\mu_{\Delta\alpha}, \tau_{\Delta\alpha}, 0) \qquad (10)$$

$$\Delta\beta \sim N(\mu_{\Delta\beta}, \tau_{\Delta\beta}, 0) \qquad (11)$$

The analysis consisted of two parts. In the first part, for each growing season, the model was fitted to the grape bunch data available after the first, second, and up to the final (fifteenth) observation time, assuming independent residuals over time, due to the destructive measuring procedure. This was done in order to investigate the trade-off between an earlier prediction and prediction accuracy, as well as to look at the growth curves fitted after one season. In the second part, four different sets of priors were taken into account when analysing the bunch mass data for the 2018/19 growing season in order to examine the effects of incorporating historical data into the modelling process. The weakly informative priors were based on the general expectations of the shape of the grape growth curve and had high variance, and thus small precision. The more informative prior distributions were obtained by using the means and variances of the posterior sample's

TABLE 1. Prior distributions used for modelling the growth of grape bunches in the 2018/19 season. The vague prior is based on general expectations, while the parameters of the 2017, 2018, and 2017 + 2018 priors are based on parametric approximations of the posterior distributions resulting from combining vague prior with the 2017 and 2018 seasons, and a combination of the two years of bunch mass data.

Coefficient	Vague Prior	2017 Prior	2018 Prior	2017+2018 Prior
α_0	N(4.09,0.11)	N(4.40,32.61)	N(4.58, 155.24)	N(4.31, 100.77)
Δα	TN(0.69,4,0)	TN(0.76,13.28,0)	TN(0.74,121.99,0)	TN(2.01,27.62,0)
β_0	N(200,0.02)	N(186.89,0.93)	N(171.26,2.97)	N(181.25,1.75)
Δβ	TN(49,0.11,0)	TN(58.48,0.05,0)	TN(53.38,0.28,0)	TN(58.61,0.05,0)
γ_0	TN(0.3,44.44,0)	TN(0.09,28591.51,0)	TN(0.20,104.75,0)	TN(0.07,94918.84,0)
γ_1	TN(0.3,44.44,0)	TN(0.12,182.66,0)	TN(0.27,90.17,0)	TN($5.19*10^{-5}$,39484085,0)
τ	Gamma(4,1)	Gamma(208.72,31.55)	Gamma(207.62,22.22)	Gamma(202.81,26.61)

parametric approximation of the posterior distributions from the analysis of the 2017/18 bunches, the 2016/17 bunches, and by combining the two. This was done by taking the informed priors from the 2017/18 season's bunches and informing these on the 2016/17 grape bunches. The three sets of priors are listed in Table 1.

A Metropolis-Hasting sampler was written in R (Team (2015)). In each case, 10^5 iterations were run after a 2×10^4 burn-in and the convergence was visually assessed.

2.3 Simulation studies

In order to examine the effect of prior assumptions on yield prediction, as well as to investigate the value of information in the context of yield prediction, 100 data sets were simulated based on the parameters drawn from the posterior distribution, produced after fitting the model to the 2018/19 season's bunch mass data. The simulated data consisted of the mass of 15 grape bunches being measured at 15 time points throughout the growing season. The model described in Equations (1)-(11) was then fitted with the four different priors described in Table 2. The estimated posterior average bunch mass at day 273 (the harvest day for the year 2019) was then evaluated after each of the 14 measurement time points. In addition, the mean absolute error (MAE) was defined as

$$MAE = \left\| E(Y \vee x = 273) - \overline{y}_{273}^{sim} \right\| \qquad (12)$$

Where

$$E\left(Y|x = 273, y^{sim}\right) = E\left\{\frac{\alpha_0}{1+e^{-\gamma_0(273-\beta_0)}} + \frac{\alpha_1}{1+e^{-\gamma_1(273-\beta_1)}}\right\}$$

Y^{sim} is the simulated data set and is the simulated average bunch mass on day 273. Similarly, the mean percentage error (MPE) was evaluated as

$$\text{MPE} = \log\frac{E(Y \vee x=273)}{\overline{y}_{273}} \qquad (13)$$

RESULTS

Individual bunch masses with the running (geometric) mean are shown in Figure 2. Interesting differences between the seasons are apparent. A clear Phase II is apparent in the 2017/18 season between days 25 and 50 (15th January to 8th February 2018), which was not observed in 2016/17. This reflects the lower average temperature over the flowering period in 2016/17 when compared to 2017/18, resulting in a longer flowering duration.

The posterior estimated mean growth curves for the grape bunches in the three seasons examined are shown in Figure 2. While all the models show a reasonable fit, there are a couple of aspects worth noting. A clear phase II and double sigmoid curve is apparent in the 2017/18 season between days 25 and 50 (15th January to 8th February 2018) and to a lesser extent in 2018/19. This was not observed in 2016/17, for which the resulting growth appears to be a simple sigmoid curve. Some of these differences may be explained by the seasonal differences in average

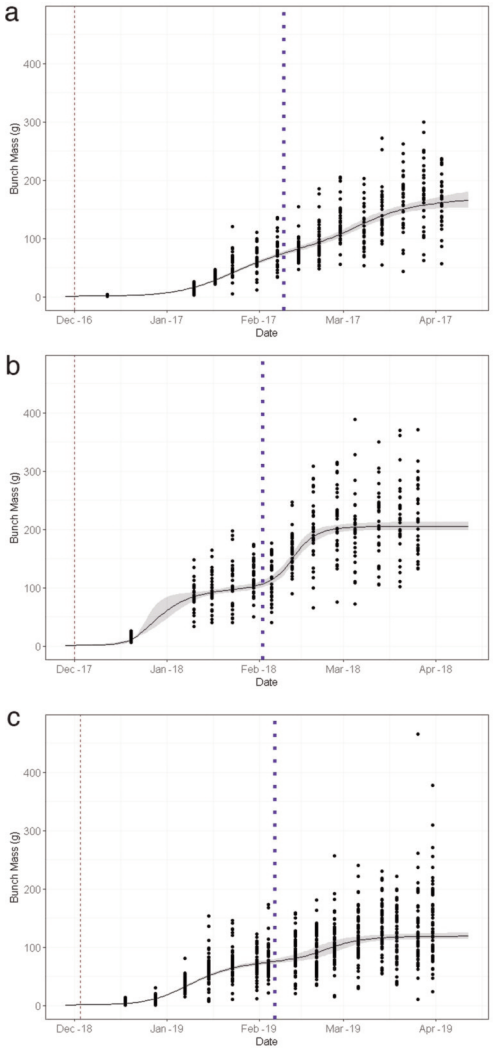

FIGURE 2. Individual observations and average bunch masses (g) (solid black line) of grape bunches over the 2016/17, 2017/18 and 2018/19 growing seasons.
The black dotted lines in each plot represent the posterior estimated mean growth curve for each season, with the surrounding grey ribbons representing the 10 % and 90 % estimation bounds. 2a. 2016/17 Growing season, 2b. 2017/18 Growing season, 2c. 2018/19 Growing season

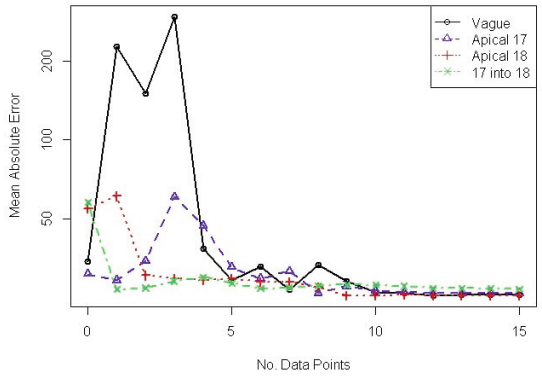

FIGURE 3. The posterior mean absolute error for the models with the vague prior, along with the 2017, 2018, and 2017+18 priors, fitted to the grape bunch masses from the 2018/19 season after different amount of observations have been made available.

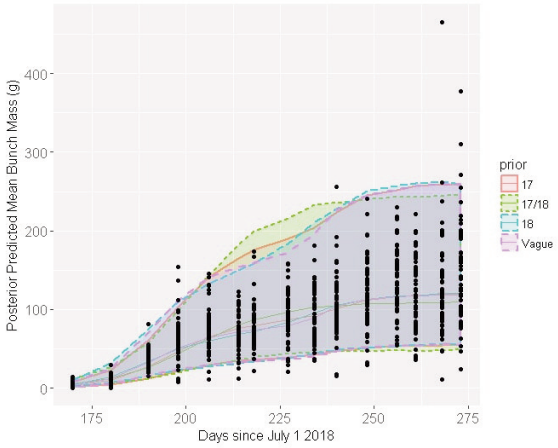

FIGURE 4. Posterior mean bunch mass curves with bounds for the double sigmoidal model fitted to the data for the 2018/19 growing season with the vague priors, along with priors informing using the 2016/17 data, the 2017/18 data, and combining them together in different orders (2016/17 to 2017/18 and vice versa). The 5 % and 95 % estimation bounds are included for each set of priors.

temperature during flowering, and in rainfall from flowering to véraison. The average daily temperature, calculated using the model described in Trought (2005), in 2016/17 was 17.9 °C, while those in 2017/18 and 2018/19 were 18.8 and 18.6 °C respectively. Lower temperatures result in a longer flowering duration and greater asynchronous berry development. The rainfall from flowering to véraison was significantly higher in 2017/18 (135 mm) than in 2016/17 and 2018/19 (49 and 41 mm respectively).

Rainfall during fruit ripening resulted in the onset of *Botrytis cinerea* leading to a sharp deterioration of bunch mass in 2017/18 (Figure 2b). The data for the last observational time point were therefore not included in the model, because the sharp decrease in bunch mass would have affected the harvest mass estimates produced from the Bayesian model for that particular year; here, the standard double sigmoidal growth curve struggles to accommodate the sharp rise in the middle of the season.

The posterior estimates for the average bunch mass were also obtained after adding the data as it became available consecutively to the model along with the various sets of priors. The resulting mean absolute errors are shown in Figure 3. The value of information is evident here, along with the importance of having informed priors, particularly at the starting of the growing season. This is seen where the prediction errors for the model with vague priors are quite poor and erratic, although they do improve markedly once five or more data points are added (23rd January 2019).

Figure 4 shows the posterior mean curves and associated 95 % credible bounds, which were estimated after fitting the model to the entire 2018/19 bunch mass data with the vague, the 2017, the 2018, and the 2017 + 2018 priors respectively. The results show that there is some apparent sensitivity to the choice of priors at different points in the growing season. For example, for the results from the incorporation of two years of priors, the estimation bounds are somewhat larger than those from the other three sets of priors. However, this switches to having lower estimations for the bunch masses at the end of the growing season.

Figure 5 compares the posterior estimates for the final bunch masses of the 2018/19 bunch mass data, using each of the three informed sets of priors (2017, 2018, and 2017 + 2018). When estimates are made using two years of data, they become more consistent earlier in the growing season, after about half of the data has been included in the Bayesian model. However, the means of these estimates appear to be lower than

the actual final yield, as well as the estimates made from the other two models (2017, 2018).

3.1 Simulation studies

For each of the 100 simulated datasets, the posterior mean bunch masses at harvest (273rd day) and the associated 95 % credible interval were evaluated *a priori* and after each of the 14 observation time points. The results are shown in Figures 6, 7, 8 and 9. The informative priors naturally have narrower credible intervals to begin with. As the season progresses and more observations are collected, the intervals become narrower and the predictions for the average

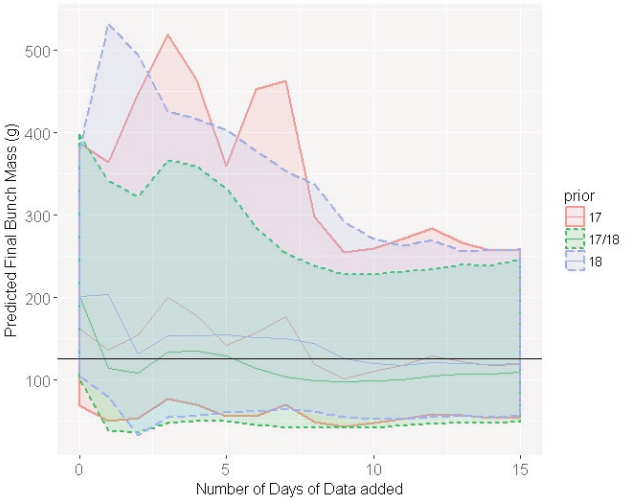

FIGURE 5. The predicted final bunch masses for the 2018/19 growing season, compared over the 3 sets of informed priors. The 5 % and 95 % estimation bounds are included for each set of priors.
The black horizontal line represents the observed average final yield for the 2018/19 growing season.

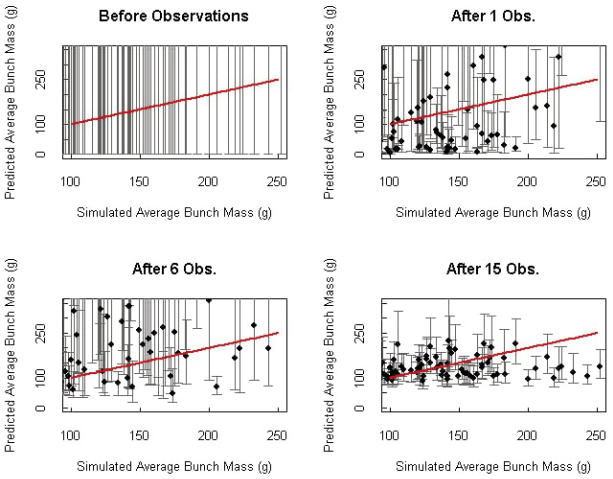

FIGURE 6. Posterior mean masses at harvest (273rd day) and the associated 95 % CIs for the 100 simulated datasets modelled using the vague prior a priori and after the 1st (18th December), 6th (31st January) and 15th (31st March) observation time points respectively.

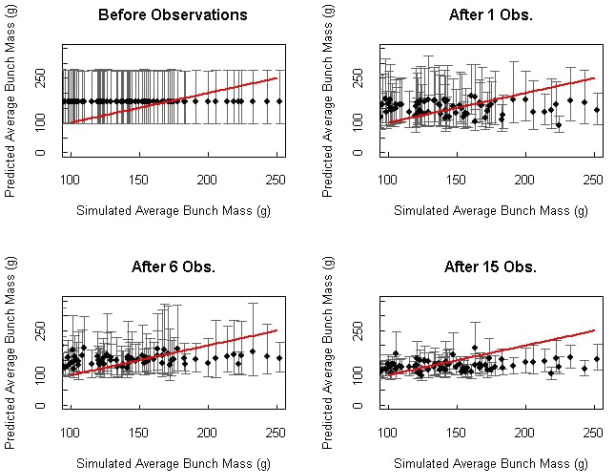

FIGURE 7. Posterior mean masses at harvest (273rd day) and the associated 95 % CIs for the 100 simulated datasets modelled using the 2017 prior a priori and after the 1st, 6th and 15th observation time points respectively.
The red line indicates equivalence.

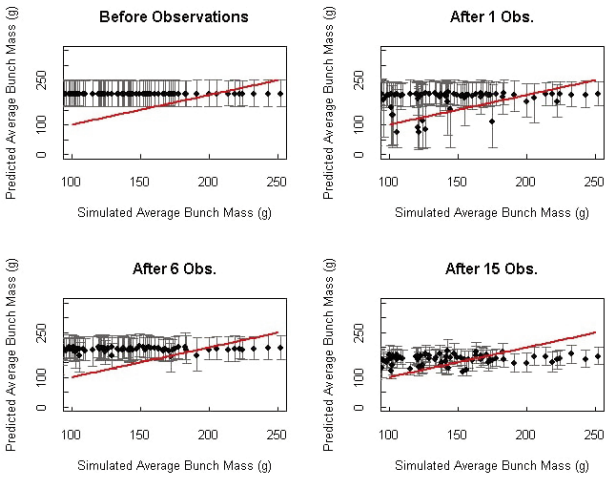

FIGURE 8. Posterior mean masses at harvest (273rd day) and the associated 95 % CIs for the 100 simulated datasets modelled using the 2018 prior a priori and after the 1st, 6th and 15th observation time points respectively.
The red line indicates equivalence.

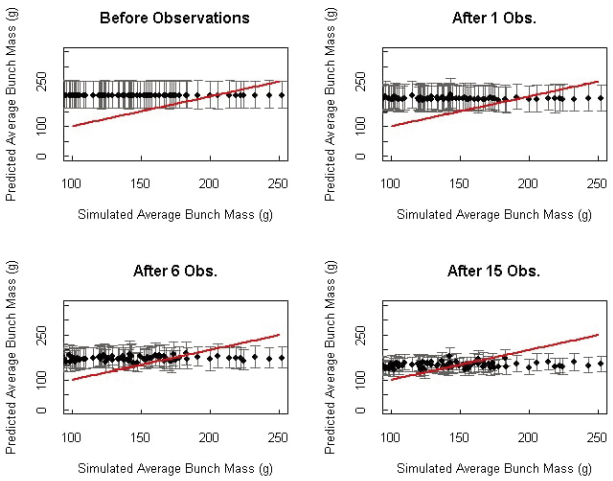

FIGURE 9. Posterior mean masses at harvest (273rd day) and the associated 95 % CIs for the 100 simulated datasets modelled using the 2017 + 2018 prior a priori and after the 1st, 6th and 15th observation time points respectively.
The red line indicates equivalence.

bunch mass at harvest become individually dependent on the dataset and closer to the black line, indicating a perfect guess. It is worth noting that the prior based on the 2017/18 growing seasons, which was very different from the 2016/17 one, is overly conservative and is not swayed by the data, resulting in relatively poor predictions, even at the end of the season. The model based on priors informed from both of these years of data suffers even more from this, as the increased precision on the asymptote parameters (α_0 and $\Delta\alpha$) keeps the means of the posterior estimates more fixed and with tighter bounds. The model with priors informed using two seasons of bunch mass data tended to overestimate the bunch masses with less data, and underestimate with all the data.

The resulting mean absolute error, averaged over the 100 datasets, is shown in Figure 10. Here, the vague prior only starts to perform just as well as the other priors once approximately 14 of the 15 days of data have been included, which is not suitable given the desire to produce a model which can produce reliable estimates early in the growing season. However, the results shown in Figures 10 and 11 suggest a high consistency throughout the growing season when using the other three priors.

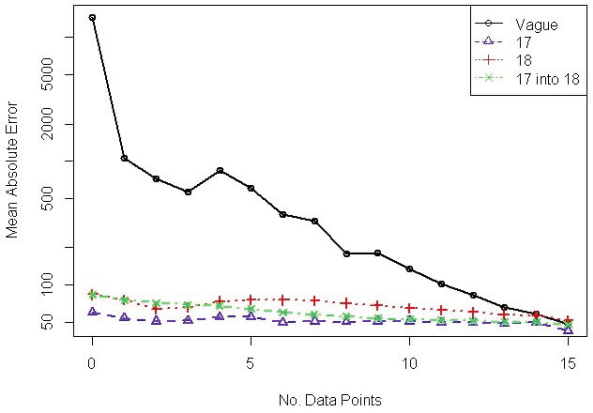

FIGURE 10. Mean absolute error for the bunch mass at harvest averaged over the 100 simulated datasets for the vague, 2017, 2018, and 2017 + 2018 priors respectively.

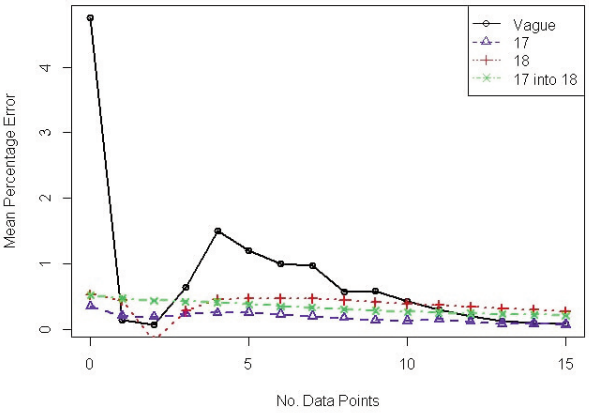

FIGURE 11. Mean percentage error for the mass bunch at harvest averaged over the 100 simulated datasets for the vague, 2017, 2018, and 2017 + 2018 priors respectively.

DISCUSSION

Bayesian methods are capable of systematically incorporating prior knowledge. This feature is especially relevant to viticulture and to grape growth modelling, since there is substantial, often vineyard-specific, expert knowledge available. The ability of the Bayesian framework to seamlessly update model estimates as new data comes in is especially useful given the dynamic nature of the phenomenon being modelled. Starting out with a yield estimate based on historical data, and perhaps a general weather forecast for the coming season, and

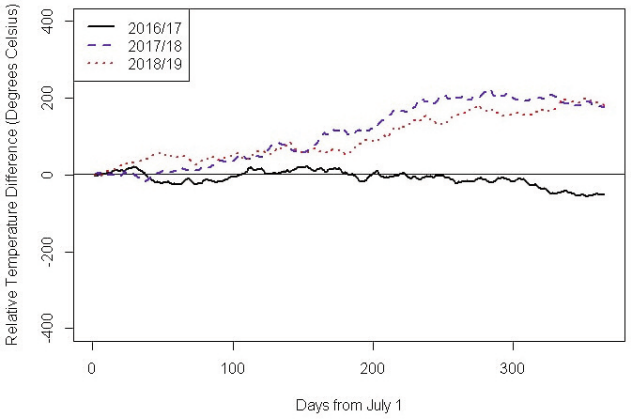

FIGURE 12. Temperature anomalies for the 2016/17, 2017/18 and 2018/19 growing seasons, relative to the average daily temperatures over the last 10 years.

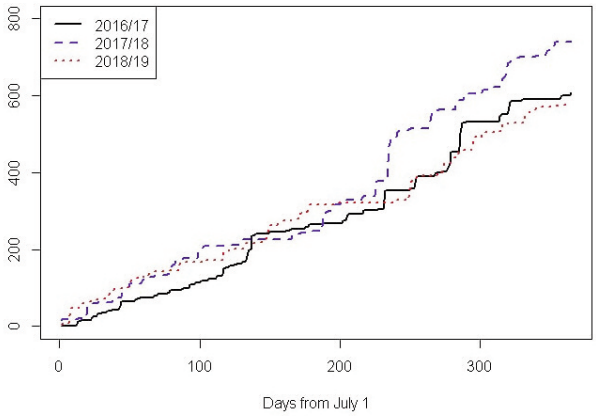

FIGURE 13. Temperature anomalies for the 2016/17, 2017/18 and 2018/19 growing seasons, relative to the average daily temperatures over the last 10 years.

revising that estimate as new information becomes available is a goal well-worth achieving.

The model examined in this study considers the bunch mass to be a function of time only. This is clearly an oversimplification, since plant growth in general, and grape growth in particular, is known to be affected by temperature, usually expressed as growing degree days (Coombe, 1986). Wang and Engel (1998) introduced a function describing the relationship between the daily temperature and the daily growth date and applied it to wheat growth, whereas, for example, Parker *et al.* (2011) and Parker (2012) extended these ideas to model phenological stages of grape development. Therefore, our model may be improved by replacing days with the growth rate expressed as a function of growing degree days. Climate variables other than temperature, such as the amount of solar radiation, may also influence grape growth and development (Dokoozlian and Kliewer, 1996; Bergqvist *et al.*, 2001; Fernandes de Oliveira and Nieddu, 2016). Additional variations may arise due to characteristics of the land, such as soil and topography (Trought and Bramley, 2011; Bramley *et al.*, 2011), as well as management practices. When available, this information can be incorporated into the framework by hierarchically modelling the growth curves parameters. Thus, for example, the inflection points β_0 and β_1 can be construed as functions of climate and spatial covariates.

As a way of emphasising this, we took into consideration climate information for the region in which the data was collected. Meteorological data was sourced from the National Institute of Water and Atmospheric Research at the Marlborough Research station 1.1 km south west of the trial site (49°21'51'S; 173° 47' 56"E). Figure 12 demonstrates the temperature anomalies for the three growing seasons relative to the last 10 years. The accumulated temperature experienced in the 2016/17 growing season was lower than in the other seasons. Above-average temperatures were noted in the 2017/18 season over the flowering period (as indicated by the slope of the deviation from the long-term mean), while the other two seasons were close to the long-term average over flowering. Further heatwaves were experienced in both the 2017/18 and 2018/19 seasons at certain times in the period between flowering and véraison (Salinger *et al.*, 2019; Salinger *et al.*, 2020). The dates of flowering and véraison were estimated using the Grapevine Flowering Véraison (GFV) (Parker *et al.*, 2011). This appeared to have more of a direct impact on the yield for the 2017/18 seasons, shown by the increase in the average bunch mass measurements. Figure 13 shows the accumulated rainfall for the three growing seasons. It can be seen that the 2017/18 growing season experienced higher amounts of rainfall during the course of the growing season. The 2018/19 season is typified by extended periods of no rainfall in the second half of the season. This may explain the very limited amount of growth

in the bunch masses seen after the period of véraison in this particular case.

It can be seen in Figure 2a that the 6-parameter double sigmoidal curve was not a good fit for the grape growth observed in the 2016/17 season. Archontoulis and Miguez (2015) have reviewed a wide variety of nonlinear regression models used in agricultural research and they have made comparisons between various specifications of sigmoid functions. Of particular interest to us are the Richards model (Richards,1959), the Gompertz model (Gompertz, 1825) and the Weibull curve (Weibull, 1951), all of which we intend to compare with our current implementation of the double sigmoidal model. It is also worth noting another more recent model developed by Yin *et al.* (2003), which involves incorporating the maximum growth rate of the fruit to help calculate fruit mass.

The current model assumed conditionally independent priors and was estimated via the standard Metropolis-Hasting sampler (Gelman *et al.*, 2014). Taking joint prior specifications into account will increase the flexibility of the model, allowing a wider range of expert opinion formulations to be adapted, and may also improve the efficiency of the algorithm.

Finally, applying the resulting range of parametrisations to a wider spatio-temporal set of data will improve our understanding of the modelling framework, and ultimately produce a better tool for early grape yield prediction.

CONCLUSION

In this study, we illustrated the use of a Bayesian framework for fitting a standard double sigmoidal growth curve to the Sauvignon blanc grape bunch mass data collected over the 2016/17 2017/18 and 2018/19 growing seasons in Marlborough, New Zealand. We also performed a simulation study to investigate both the sensitivity of the model to prior assumptions and the value of information. The latter refers to the role of additional consecutive observations throughout the season in improving the accuracy of the estimation of the grape bunch mass at harvest.

The results from this analysis show that the model is sensitive to prior assumptions made for the parameters of the double sigmoidal model. From an early yield prediction perspective, the incorporation of non-informative (vague) priors to the model resulted in poor results, only becoming similar to the results seen from models using more informed priors once half of the growing season data was incorporated into the Bayesian model. This led to the conclusion that some information about the parameterisation of the double sigmoidal model is influential in producing useful results.

Acknowledgements: We would like to thank the Marlborough Research Centre Trust for the use of their vineyard.

REFERENCES

Archontoulis S.V. and Miguez F.E., 2015. Nonlinear regression models and applications in agricultural research. *Agronomy Journal* **107(2)**, 786-798. https://doi.org/10.2134/agronj2012.0506

Bergqvist J., Dokoozlian N. and Ebisuda N., 2001. Sunlight exposure and temperature effects on berry growth and composition of cabernet sauvignon and grenache in the central San Joaquin valley of California. *American Journal of Enology and Viticulture* **52(1)**, 1-7.

Bramley R.G., Trought M.C.T. and Praat J.-P., 2011. Vineyard variability in Marlborough, New Zealand: characterising variation in vineyard performance and options for the implementation of precision viticulture. *Australian Journal of Grape and Wine Research* **17(1)**, 72-78. https://doi.org/10.1111/j.1755-0238.2010.00119.x

Coombe B., 1976. The development of fleshy fruits. *Annual Review of Plant Physiology* **27(1)**, 207-228. https://doi.org/10.1146/annurev.pp.27.060176.001231

Coombe B., 1986. Influence of temperature on composition and quality of grapes. *Symposium on Grapevine Canopy and Vigor Management*, XXII IHC 206, pp. 23-36.

Coombe B., 1992. Research on development and ripening of the grape berry. American *Journal of Enology and Viticulture* **43(1)**, 101-110.

Coombe B.G. and McCarthy M., 2000. Dynamics of grape berry growth and physiology of ripening. *Australian Journal of Grape and Wine Research* **6(2)**, 131-135. https://doi.org/10.1111/j.1755-0238.2000.tb00171.x

Dokoozlian N. and Kliewer W., 1996. Influence of light on grape berry growth and composition varies during fruit development. *Journal of the American Society for Horticultural Science* **121(5)**, 869-874. https://doi.org/10.21273/JASHS.121.5.869

Fernandes de Oliveira A. and Nieddu G., 2016. Vine growth and physiological performance of two red grape cultivars under natural and reduced UV solar radiation. *Australian journal of grape and wine*

research **22(1)**, 105-114. https://doi.org/10.1111/ajgw.12179

Fernandes T.J., Pereira A.A. and Muniz J.A., 2017. Double sigmoidal models describing the growth of coffee berries. *Ciencia Rural* **47(8)**. https://doi.org/10.1590/0103-8478cr20160646

Galassi A., Cappellini P. and Miotto G., 2000. A descriptive model for peach fruit growth. *Advances in Horticultural Science* 19-22.

Gelman A., Carlin J. B., Stern H. S., Dunson D. B., Vehtari A. and Rubin D. B., 2014. *Bayesian Data Analysis*, Third Edition (CRC Press, Taylor and Francis: Boca Raton). https://doi.org/10.1201/b16018

Gompertz B., 1825. Xxiv. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS andc. *Philosophical transactions of the Royal Society of London* **115**, 513-583. https://doi.org/10.1098/rstl.1825.0026

Mun J., 2002. *Real options analysis: Tools and techniques for valuing strategic investments and decisions*. Volume 137 - John Wiley and Sons.

Parker A., De Cortázar-Atauri I.G., van Leeuwen, C. and Chuine I., 2011. General phenological model to characterise the timing of flowering and veraison of *Vitis vinifera l.*. *Australian Journal of Grape and Wine Research* **17(2)**, 206-216. https://doi.org/10.1111/j.1755-0238.2011.00140.x

Parker A.K., 2012. Modelling phenology and maturation of the grapevine *Vitis vinifera* L.: varietal differences and the role of leaf area to fruit mass ratio manipulations*. PhD thesis*, Lincoln University.

Parra-Coronado A., Fischer G. and Camacho-Tamayo J.H., 2016. *Growth model of the pineapple guava fruit as a function of thermal time and altitude*. Ingeniería e Investigación **36(3)**, 6-14. https://doi.org/10.15446/ing.investig.v36n3.52336

R Core Team, 2013. *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Richards F., 1959. A flexible growth function for empirical use. *Journal of experimental Botany* **10(2)**, 290-301. https://doi.org/10.1093/jxb/10.2.290

Salinger M.J., J Diamond H.J., Behrens E., Fernandez D., Fitzharris B.B., Herold N., Johnstone P., Kerckhoffs H., Mullan A.B., Parker A.K., Renwick J., Schofield, C., Siano A., Smith R.O., South P.M., Sutton P.J., Teixeira E., Thomsen M.S. and Trought M.C.T. , 2020. Unparalleled coupled ocean-atmosphere summer heatwaves in the New Zealand region: Drivers, mechanisms and impacts. Climate Change (in press). https://doi.org/10.1007/s10584-020-02730-5

Salinger M.J., Renwick J., Behrens E., Mullan A.B., Diamond H.J., Sirguey P., Smith R.O., Trought M.C.T., Alexander L.V., Cullen N.J., Fitzharris B.F., Hepburn, C.D., Parker A.K. and Sutton P.J., 2019. The unprecedented coupled ocean-atmosphere summer heatwave in the New Zealand region 2017/18; drivers, mechanisms and impacts. *Environmental Research Letters* 14, 044023. https://doi.org/10.10.1088/1748-9326/ab012a. https://doi.org/10.1088/1748-9326/ab012a

Salisbury F.B. and Ross C., 2000. *Fisiología de las Plantas 3. Desarrollo de las plantas y fisiología ambiental* (Spain, Paraninfo S.A.: Thompson Editores).

Tarara J.M., Blom P.E., Shafii B., Price W.J. and Olmstead M.A., 2009. Modeling seasonal dynamics of canopy and fruit growth in grapevine for application in trellis tension monitoring. *HortScience* **44(2)**, 334-340. https://doi.org/10.21273/HORTSCI.44.2.334

Trought M.C.T., 2005. *Fruitset - possible implications on wine quality*. Garis K. d., Dundon,C., Johnstone R. and Partridge S., ed. Transforming flowers to fruit; Mildura, Australia. *Australian Society of Viticulture and Oenology*, 32-36.

Trought M.C.T. and Bramley R.G., 2011. Vineyard variability in Marlborough, New Zealand: characterising spatial and temporal changes in fruit composition and juice quality in the vineyard. *Australian Journal of Grape and Wine Research* **17(1)**, 79-89. https://doi.org/10.1111/j.1755-0238.2010.00120.x

Wang E. and Engel T., 1998. Simulation of phenological development of wheat crops. *Agricultural systems* **58(1)**, 1-24. https://doi.org/10.1016/S0308-521X(98)00028-6

Weibull W., 1951. A statistical distribution function of wide applicability. Journal of Applied Mechanics-Transactions of the ASME **19(2)**, 293-297.

Yin X., Goudriaan J., Lantinga E. A., Vos J. and Spiertz H. J., 2003. A flexible sigmoid function of determinate growth. *Annals of Botany* **91(3)**, 361-371. https://doi.org/10.1093/aob/mcg029

